

UNIVERSIDADE DE LISBOA

Faculdade de Medicina



Parâmetros populacionais e forenses da população do Sul de Portugal

Atualização dos dados genéticos dos loci STR usados na casuística forense

Raquel Cabezas da Silva

Orientadores: Mestre Paulo Alexandre Paisana da Silva Dario

Professor Doutor Jorge Manuel Matias da Costa Santos

Dissertação especialmente elaborada para obtenção do grau de
Mestre em Medicina Legal e Ciências Forenses

2016

A impressão desta dissertação foi aprovada pelo Conselho Científico da Faculdade de Medicina de Lisboa em reunião de 19 de Janeiro de 2016.

*O ADN é a chave mestra da vida,
E ainda restam muitas portas por abrir
Para por nós ser lida*

-Raquel C. Silva

Errata

Errata referente à dissertação de Mestrado intitulada “Parâmetros populacionais e forenses da população do Sul de Portugal - Atualização dos dados genéticos dos *loci* STR usados na casuística forense”, realizada por Raquel Cabezas da Silva.

Página	Linha	Onde se lê	Leia-se
23	1 e 2	Do conselho de 25 de Junho de 2001 da União Europeia	de 25 de Junho de 2001 do Conselho da União Europeia
23	8	conselho	Conselho
25	14	Penta D e Penta D	Penta D e Penta E
43	11	Ribeiro et al.	Vieira-Silva

Agradecimentos

Ao Professor Doutor Jorge Costa Santos pela sua disponibilidade imediata em coorientar este trabalho.

Ao Mestre Paulo Dario não só na qualidade de orientador mas também por todo o seu apoio, atenção e motivação. Um especial obrigado por todos os conhecimentos e sabedoria que me transmitiu ao longo deste trabalho.

À Dra. Teresa Ribeiro por me conceder a oportunidade de estagiar no laboratório, onde tive o privilégio de aprender com os melhores profissionais.

Aos demais membros do serviço de Genética e Biologia Forense pela ajuda disponibilizada e por todos os conhecimentos transmitidos, em especial à Mestre Rita Dario e à Dra. Isabel Lucas por toda a ajuda dentro e fora do laboratório.

À Dra. Manuela Marques por toda a sua disponibilidade, eficiência e apoio na minha pesquisa bibliográfica.

Aos meus amigos João Bessa Ferreira, Isa Pinto e Tara Canevascini pela sólida e duradoura amizade.

Aos meus pais, Vicente e Isabel por todo o seu amor, apoio e incentivo. Sem eles este trabalho jamais seria possível.

Lista de abreviaturas

A – Adenina

ADN- Ácido desoxirribonucleico

C - Citosina

CODIS - *Combined DNA Index System*

DIP - *Deletion or Insertion Polymorphisms*, polimorfismos de deleção ou inserção

EDNAP - *European DNA Profiling Group*

F_{ST} - Índice de fixação

G - Guanina

H - Heterozigotia

He - Heterozigotia esperada

Ho - Heterozigotia observada

HWE – *Hardy-Weinberg equilibrium*, equilíbrio de Hardy-Weinberg

INMLCF,I.P - Instituto Nacional de Medicina Legal e Ciências Forenses, Instituto Público

INTERPOL - *International Criminal Police Organization*

IP - Índice de paternidade

ISFG - *International Society for Forensic Genetics*

LR - *Likelihood Ratio*, razão de verosimilhança

MAF - *Minimum Allele Frequency*, frequência do alelo mínimo

MEGA - *Molecular Evolutionary Genetics Analysis*

NDNAD - *National DNA Database*

Pb – pares de bases

PCR - *Polymerase Chain Reaction*, reação em cadeia da polimerase

PD - Poder de Discriminação

PE - Poder de Exclusão

PIC - *Polymorphic Information Content*, conteúdo de informação polimórfica

SGBF – Serviço de Genética e Biologia Forenses

SNP - *Single Nucleotide Polymorphism*, polimorfismo de base única

STR - *Short Tandem Repeat*, pequena repetição concatenada

T - Timina

TPI - *Typical Paternity Index*, índice de paternidade típico

VNTRs - *Variable Number of Tandem Repeats*, número variável de repetições concatenadas

Índice

Agradecimentos.....	1
Lista de Abreviaturas.....	2
1. Resumo/Abstract.....	7
2. Introdução.....	12
2.1. Breve história da Genética Forense.....	13
2.2. Marcadores genéticos.....	18
2.2.1. <i>Short tandem repeats</i>	19
2.2.2. Conjunto de marcadores obrigatórios.....	21
2.2.3. <i>Kits</i> comerciais.....	24
2.2.4. Bases de dados de ADN.....	26
2.3. Genética Populacional e Estatística em contexto legal.....	30
2.3.1. Genética Populacional.....	30
2.3.2. Parâmetros estatísticos de interesse forense.....	34
2.3.3. Valor estatístico do ADN.....	38
3. Objetivos do Estudo.....	42
4. Material e Métodos.....	45
4.1. Rotina laboratorial.....	46
4.2. Amostragem e compilação dos dados.....	48

5. Resultados e Discussão.....	51
5.1. Frequências alélicas.....	52
5.2. Parâmetros estatísticos de interesse forense.....	57
5.3. Comparação interpopulacional.....	59
5.4. Comparação com o estudo de 2006.....	61
 6. Conclusões.....	 62
 7. Referências Bibliográficas.....	 64
 8. Anexos.....	 70

1. Resumo/Abstract

O objetivo da Genética Forense prende-se com a identificação, seja no âmbito da criminalística biológica, investigações de parentesco ou identificação individual. Perante um caso de não-exclusão é apresentada uma probabilidade de identidade. Este cálculo é realizado usando as frequências alélicas de cada alelo que um determinado perfil genético reúne.

Este trabalho teve como finalidade a atualização das frequências alélicas e dos parâmetros estatísticos com interesse forense da população abrangida pelo Serviço de Genética e Biologia Forenses da Delegação do Sul do Instituto Nacional de Medicina Legal e Ciências Forenses, I.P. Para concretizar o estudo foram utilizados dados genéticos de 5362 indivíduos portugueses caucasianos não aparentados envolvidos em investigações de parentesco biológico, entre 2005 e 2014. Foi estudado um conjunto de 17 *loci* STR incluídos nas famílias dos kits *AmpFlSTR Identifiler™* e *Powerplex® 16 System* utilizados na rotina do laboratório, nomeadamente: CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, FGA, TH01, TPOX, vWA, D2S1338, D19S433, Penta D e Penta E. O último estudo semelhante com estes marcadores foi realizado em 2006, pelo que as frequências podem ter sofrido alterações ao longo destes quase dez anos, em consequência de fenómenos evolutivos da população.

Foram calculadas as frequências alélicas e os parâmetros estatísticos de interesse forense como a heterozigotia observada, o poder de discriminação, o poder de exclusão, o conteúdo de informação polimórfica e o índice de paternidade típico utilizando a folha de cálculo de Excel Powerstats v.1.2 modificada. A heterozigotia esperada, o equilíbrio de Hardy-Weinberg e os valores de F_{ST} foram determinados através do *software* Arlequin v.3.5.

De todo o conjunto de marcadores, foi demonstrado que o marcador TPOX é o menos polimórfico enquanto o Penta E mostrou ser o mais polimórfico. Todos os marcadores

apresentaram um valor de heterozigotia esperada superior a 85%. O poder de discriminação acumulado corresponde a 0,99999999999999999997 para o conjunto dos 17 marcadores.

Foram também comparadas as frequências alélicas entre as populações do sul de Portugal, Espanha, Itália, Grécia, Roménia, Marrocos, Angola e Coreia. A análise filogenética mostrou que as populações geneticamente mais afastadas de Portugal são Angola e a Coreia, com um valor de F_{ST} de 0,0111 e 0,0197, respetivamente.

De uma forma global, as frequências alélicas para os alelos presentes no estudo anterior não sofreram alterações muito significativas. No entanto, detetou-se a presença de um número mais elevado de variantes alélicas, evidenciando-se um total de 81 alelos diferentes para os 17 *loci* STRs, enquanto no estudo anterior havia um total de 66 alelos, o que significa um incremento de cerca de 22%. Este aumento pode estar fortemente relacionado com o tamanho da amostragem, pelo que se conclui que quanto maior for o número de amostras, mais realista e objetiva será a representação de todos os alelos presentes na população.

Palavras-chave:

Genética forense, frequências alélicas, STRs, marcadores genéticos, sul de Portugal

The objective of Forensic Genetics is the achievement of identification in its three areas: criminal investigation, kinship testing and individual identification. In case of matching it is required to present the probability of identity. This calculation is accomplished by using allele frequencies for the alleles that a DNA profile presents.

This work intended to update allele frequencies and other relevant forensic statistical parameters of the population covered by the South Branch of the National Institute of Legal Medicine and Forensic Sciences. For this study, it was used genetic data from 5362 unrelated Caucasian Portuguese individuals involved in paternity testing casework since 2005 to 2014. A set of 17 STR loci included in the AmpFISTR Identifiler™ and Powerplex® 16 System kits were studied, namely: CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, FGA, TH01, TPOX, vWA, D2S1338, D19S433, Penta D and Penta E. The last study covering this problematic dates to 2006 and it might be possible that the frequencies undergone some changes due to population evolution over the last ten years.

Allele frequencies and statistical parameters of forensic interest such as observed heterozygosity, power of discrimination, power of exclusion, polymorphic information content, typical paternity index and matching probability were calculated with modified PowerStats v1.2. Expected heterozygosity, Hardy-Weinberg equilibrium and F_{ST} values were calculated using Arlequin v.3.5 software.

From the entire set of markers, TPOX proved to be the least polymorphic marker and Penta E the most polymorphic marker. All the markers showed an expected heterozygosity value greater than 85%. The combined power of discrimination equals 0.99999999999999999997 for the whole set.

A comparison was made between allele frequencies from the southern Portuguese population and from Spain, Italy, Greece, Romania, Morocco, Angola and Korea. Phylogenetic analysis showed populations from Angola and Korea having the greatest genetic distance from southern Portugal's population with F_{ST} values of 0.0111 and 0.0197, respectively.

In general, the allele frequencies for both present and past studies do not display great differences. However, a wide number of new alleles are present. A total of 81 alleles contrasting to 66 in the previous study, results in an increase of 22% on the number of alleles. This increment might be highly related with sampling. Once the greater the number of samples, the more probable will be the representation of all the alleles present in the population.

Keywords:

Forensic genetics, allele frequencies, STRs, genetic markers, south Portugal

2.Introdução

“Karkov teria que os envenenar para lhes destruir a identidade russa, antes de deixar o hotel. Mortos, ninguém poderia provar a identidade dos três homens, um com três balas no abdômen, outro com o queixo rebentado e as cordas vocais à mostra e o terceiro com o fêmur esmagado e as mãos e a face tão queimadas que o rosto era uma chaga sem cílios, sem sobrancelhas e sem cabelos.”

Por quem os sinos doam, Ernest Hemingway, 1940

2.1. Breve história da Genética Forense

A identificação genética iniciou-se após Karl Landsteiner ter descoberto os grupos sanguíneos e Emil von Dungern e Ludwig Hirschfeld terem comprovado a transmissão hereditária dos mesmos (Geserick 2012). Contudo, apesar de efetivos para alguns casos, os testes serológicos não apresentavam um elevado poder de discriminação.

Em 1984, Alec Jeffreys e os seus colaboradores descobriram que em certas regiões do ADN existiam inúmeras sequências repetidas que variavam em comprimento e que o número dessas repetições eram únicas em cada indivíduo. Esta variação no número de repetições concatenadas, VNTRs (do inglês *Variable Number of Tandem Repeats*), veio assim fornecer um meio de prova da unicidade de todos os indivíduos (Jeffreys 1979; Tilstone 2006). Assim, através desta análise tornou-se possível a realização de novos testes de identificação genéticos, o *DNA fingerprinting*. Este termo trata-se de uma analogia a uma impressão digital, mas a nível genético (Gill 1985).

Tal como as impressões digitais, que à primeira vista são semelhantes, a partir do momento que são analisadas, pormenorizadamente, averigua-se a sua exclusividade para cada indivíduo. E algo semelhante acontecia com a molécula de ADN. Apesar do material genético

ser igual em todos os indivíduos, quando analisado de modo minucioso é possível encontrar pequenos fragmentos conferentes de singularidade.

A técnica de identificação *DNA fingerprinting* foi utilizada pela primeira vez em 1985 num caso de imigração com vista à confirmação da identidade de uma criança britânica pertencente a uma família originária do Gana, residente no Reino Unido. Esta criança esteve algum tempo no Gana e regressou novamente ao Reino Unido. De regresso, as autoridades desconfiaram que, apesar de ser portadora de um passaporte britânico, pudesse não ser quem alegava, e que poderia tratar-se de um primo ou outro parente a tentar entrar ilegalmente no país. Para resolver a situação, as autoridades pretendiam saber se a criança era efetivamente quem dizia ser, pelo que era necessário averiguar se era filho dos pais que constavam no documento de identificação, e não outro familiar. O caso foi resolvido com sucesso, com a análise do ADN que evidenciou a filiação do menino britânico de origem ganesa (Jeffreys 1985).

Passado pouco tempo a mesma técnica de identificação foi utilizada para resolver, com êxito, um caso de duplo homicídio em Leicestershire, Inglaterra. Duas raparigas, ambas com quinze anos, foram violadas e assassinadas, com um intervalo de três anos entre as duas mortes. Um homem confessou o segundo homicídio, mas não o primeiro. Apesar desta confissão, a polícia acreditava que ambas as mortes teriam sido provocadas pelo mesmo autor, dadas as semelhanças entre os dois casos, pelo que pediram a colaboração de Alec Jeffreys. Através da técnica de *DNA fingerprinting* foi confirmado que o sêmen encontrado nos dois corpos correspondia ao mesmo indivíduo, e que o confessor do segundo crime não era realmente o autor. A polícia iniciou então uma busca ao assassino, recolhendo amostras de todos os homens da região, não encontrando correspondência alguma com o perfil das amostras de sêmen. Passado um ano, uma mulher ouviu num bar o relato de um indivíduo que alegava ter doado uma amostra de sangue na vez de um amigo chamado Colin Pitchfork. A polícia

entrevistou Pitchfork, recolheu uma amostra e foi constatado que o seu perfil correspondia ao perfil das amostras de sémen encontradas nos locais do crime. Este indivíduo foi julgado e condenado a prisão perpétua (Wong 1987). A primeira aplicação do *DNA fingerprinting* à área forense, provou ser de grande valor porque, para além de identificar e condenar um criminoso, foi capaz de absolver uma pessoa inocente.

Desde então, o método original de *DNA fingerprinting* sofreu várias modificações e inovações, tornando a obtenção de resultados consideravelmente mais rápida, sendo atualmente designada por *DNA typing* ou *DNA profiling*, eliminando assim a analogia com a análise das impressões digitais (Jackson 2011).

Atualmente, pode definir-se que a Genética Forense consiste na aplicação de técnicas de genética molecular com o intuito de auxiliar a resolução de problemas de âmbito legal. O seu objetivo primordial é a identificação genética, seja para fins de identificação individual, de investigação de parentesco ou de criminalística biológica. Em qualquer uma das situações, a investigação apresenta-se, *grosso modo*, em duas fases distintas. A primeira etapa desenvolve-se ao nível laboratorial, onde se realiza a extração, amplificação e análise do ADN, com o fim de estabelecer o perfil genético da amostra, obtendo assim o genótipo para um dado indivíduo, dador da amostra. A segunda etapa corresponde à interpretação dos dados obtidos. Ora, esta interpretação reverte-se de um carácter puramente objetivo e imparcial, assentando em cálculos probabilísticos baseados no conhecimento *a priori* das frequências alélicas da população onde se insere a amostra. Assim sendo, conclui-se que todo e qualquer procedimento laboratorial executado num laboratório de genética forense tem exclusiva e unicamente a finalidade e o propósito da identificação.

A identificação corresponde ao reconhecimento do conjunto de características particulares que identificam um indivíduo, representando estas características o seu perfil genético, o qual será sinónimo da sua singularidade. De modo sumário, pode afirmar-se que o ADN tem a valência de individualizar qualquer indivíduo.

Além da identificação genética, cumprida pela Genética Forense, existem outras disciplinas do grande conjunto que constituem as ciências forenses cujo objetivo é a identificação, como a antropologia ou a dactiloscopia. Contudo, se um esqueleto se apresentar incompleto, no caso da antropologia, ou um cadáver não apresentar membros, no caso da dactiloscopia, a única forma de identificação possível será realizada a partir do ADN, obtido a partir dos restos cadavéricos, da qual apenas a Genética Forense se poderá encarregar, colmatando assim alguns obstáculos relativos à identificação e individualização deste tipo de análises. Isto torna-se particularmente útil em situações de desastres de massa e identificação de restos cadavéricos em adiantado estado de decomposição, onde a identificação só poderá ser obtida única e exclusivamente com base numa identificação genética, pois esta, apesar de ser mais dispendiosa fornece um resultado quantitativo relativamente à probabilidade de dois indivíduos possuírem a mesma informação genética numa determinada população (Schmitt 2006) e também é consideravelmente mais rápida. Consequentemente, e tendo em conta a comparação com outras técnicas as vantagens da utilização do ADN podem estruturar-se sob quatro critérios: quantidade e qualidade da amostra, alta probabilidade de discriminação e presença em praticamente todas as células (Calabuig 2004).

No que concerne à quantidade e qualidade da amostra, esta pode ser escassa e degradada e ainda assim proporcionar resultados, graças ao uso da técnica de PCR (do inglês *Polymerase Chain Reaction*) e análise de pequenas porções de ADN. Podem ser obtidos resultados mesmo

em material biológico exceccionalmente antigo, com centenas a milhares de anos (Gill 1994; Haak 2010; Der Sarkissian 2013)

Relativamente ao poder de discriminação, podem ser obtidas probabilidades de identidade extremamente elevadas, acima dos 99%. A probabilidade de discriminação será proporcional à análise dos *loci*, isto é, quanto maior for o número de *loci* analisados, maior será essa probabilidade.

No que respeita à presença de ADN, sabe-se que esta molécula está presente em todas as células nucleadas do corpo humano, à exceção das células anucleadas como são exemplo os eritrócitos e corneócitos, já que estas células não se dividem e têm um ciclo de vida muito breve (Busam 2010).

Conclui-se assim que a obtenção da identificação genética através do estudo de regiões hipervariáveis do ADN é uma das ferramentas mais poderosas utilizadas nas Ciências Forenses e amplamente aplicada pelo facto do ADN ser uma molécula relativamente robusta e única para cada indivíduo. A todas estas vantagens, alia-se ainda o facto das técnicas empregues serem relativamente rápidas na obtenção de resultados.

2.2. Marcadores genéticos

O genótipo de cada pessoa é formado durante a concepção e dificilmente sofrerá alterações. Apesar da aparência, os hábitos e os comportamentos poderem alterar-se ao longo da vida, o genótipo raramente se altera (Jackson 2011). Todas as pessoas são geneticamente diferentes, com exceção dos gémeos monozigóticos, apesar de mesmo estes poderem ter algumas diferenças não genéticas influenciadas pelo ambiente, e assim poderem ser distinguidos. Ora, se todos os seres humanos fossem geneticamente iguais, a identificação individual não seria possível, pois a cadeia de ADN seria exatamente igual em todas as pessoas. Afortunadamente, isto não acontece. Todos os indivíduos são geneticamente únicos tornando o *DNA typing* uma ferramenta de grande utilidade e valor científico.

Atualmente sabe-se que aproximadamente 99,7% do genoma humano é igual em todos os indivíduos. Uma vez que são os restantes 0,3% que tornam cada pessoa única, será nesta porção que se fará a pesquisa das regiões que diferem entre indivíduos. Existem inúmeras sequências de ADN repetitivas espalhadas por todo o genoma humano, localizando-se normalmente entre os genes, podendo variar em tamanho de pessoa para pessoa sem que isso tenha qualquer influência a nível genético (Butler 2010). Graças a esta grande variabilidade só será necessário analisar uma fração limitada de ADN para, sem embargo, permitir a identificação de indivíduos com uma elevada probabilidade, garantindo uma identificação estatisticamente significativa que constitua prova irrefutável em tribunal.

O genoma humano está repleto de sequências repetitivas de ADN, apresentando-se estas em todos os tamanhos de pares de bases (pb). A designação das sequências tem em consideração não só o tamanho da unidade de repetição mas, também, o número de repetições contínuas dessa sequência (Pinheiro 2009; Butler 2010). Consoante o número de nucleótidos

constituintes das unidades de repetição podemos estar perante minissatélites ou microssatélites. Os minissatélites ou VNTRs possuem unidades de repetição contendo entre 8 a 100 pb de extensão enquanto os microssatélites ou STRs (*Short Tandem Repeats*) contêm unidades de repetição entre 2 a 7 pb de extensão (Butler 2010; Butler 2012). Alguns exemplos de outros marcadores com vista à identificação utilizados em Genética Forense são os SNPs (do inglês *Single Nucleotide Polymorphisms*) e os DIPs (do inglês *Deletion or Insertion Polymorphisms*).

Com vista à adequação de um marcador genético à prática forense, este deve ostentar um conjunto de características aptas para tal prática. De entre estes atributos salienta-se o elevado polimorfismo, com taxas de mutação relativamente baixas, e a reprodutibilidade das técnicas (Dettmeyer et al. 2014).

2.2.1. Short Tandem Repeats

As regiões de ADN com unidades de repetição entre 2 a 7 pb são designadas por STRs ou microssatélites, sendo os tetranucleotídicos, ou seja, os com 4 pb, os mais utilizados. Os microssatélites localizam-se nas regiões não codificantes do genoma, nomeadamente em sequências intergénicas e intrões. Por estarem presentes em todos os cromossomas, podem ser designados de STRs autossómicos, do cromossoma X ou do cromossoma Y, consoante a sua localização.

Historicamente, a utilidade e eficiência dos STRs como instrumento de identificação humana foi constatada no início dos anos 90 com a observação da sua elevada frequência no genoma e

do seu elevado grau de polimorfismo, podendo o número e tipo de repetições ser muito variável, sendo assim muito discriminativos entre indivíduos não relacionados e entre familiares (Edwards 1991; Edwards 1992). O polimorfismo dos STRs reside no número de repetições, no tamanho da unidade de repetição e no próprio padrão. Características como a fácil amplificação por PCR *multiplex*, a obtenção de resultados em amostras degradadas e/ou exíguas e a obtenção de perfis genéticos de fácil interpretação e comparação faz dos STRs os marcadores de eleição na Genética Forense (Butler 2005).

Até 1997, apesar dos microssatélites já estarem em uso em vários laboratórios a nível mundial, não existia um consenso relativamente à nomenclatura destes na comunidade forense, o que dificultava comparações e troca de dados entre laboratórios. Foi neste ano que a ISFG (*International Society for Forensic Genetics*) publicou as regras no que respeita à nomenclatura dos STRs, as quais deveriam ser utilizadas por todos os laboratórios. Por exemplo, para a seguinte sequência de ADN:

5' –TTTCCC **TCAT TCAT TCAT TCAT TCAT TCAT** TCACCATGGA – 3'

3' – AAAGGG AGTA AGTA AGTA AGTA AGTA AGTA AGTGGTACCT – 5'

a leitura deve ser efetuada da extremidade 5' para a 3' na cadeia superior. O motivo de repetição neste exemplo é [TCAT], encontrando-se repetido seis vezes, pelo que o nome do alelo será 6. No entanto, há que ter em conta que certos alelos não possuem unidades de repetição completas, como é o caso do alelo 9.3 para o marcador TH01. Isto significa que o alelo tem 9 unidades de repetição mais três nucleótidos (Butler 2005). Se o alelo tivesse 9 unidades de repetição mais 4 nucleótidos, a designação do alelo já seria 10, dado estarmos perante um marcador tetranucleotídico.

O estabelecimento desta nomenclatura comum para todos os laboratórios a nível mundial facilitou a reprodutibilidade, a comparação e a troca de dados entre laboratórios. Assim, e tendo em consideração que neste momento as bases de dados genéticos aumentam continuamente em todo o mundo utilizando os STRs, não se prevê o seu desuso num futuro próximo (Butler 2010).

2.2.2. Conjunto de marcadores obrigatórios

A introdução da análise do ADN na Genética Forense não chegou de forma igual e homóloga a todos os laboratórios, com o vasto leque de marcadores para análise à disposição. Consequentemente, cada laboratório forense começou a usar determinados sistemas de marcadores como rotina laboratorial, tornando extremamente difícil uma eventual troca de resultados.

Deste modo, estipulou-se um conjunto de microssatélites de análise obrigatória em todos os laboratórios com vista à standardização de metodologias para que pudesse existir partilha e comparação de informação entre laboratórios a nível mundial. Este acontecimento também facilitou a criação e funcionamento das bases de dados nacionais. Nos Estados Unidos da América, o conjunto de marcadores escolhidos é designado por CODIS (*Combined DNA Index System*). Na Europa, é empregue o ESS (*European Standard Set*) (Butler 2006).

O sistema CODIS

Em 1997, foi anunciada a criação do CODIS e quais os marcadores que este sistema deveria integrar. Este conjunto é constituído por treze marcadores: CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51 e D21S11, para além da análise do marcador de género amelogenina. Já no final dos anos 90, alguns *kits* comerciais começaram a integrar este conjunto de marcadores entre outros adicionais (Budowle 1998).

European Standard Set

Na Europa o conjunto de marcadores a serem incluídos já sofreu algumas alterações relativamente ao número de marcadores a serem incluídos.

Em 1992, após acordo unânime dos membros do *European DNA Profiling Group* (EDNAP), estabeleceu-se que o caminho a seguir seria o uso de STRs devido a todas as vantagens inerentes às suas características. Com o acordo de todos os membros deste grupo, foram estabelecidos dois marcadores de análise obrigatória em todos os laboratórios participantes: o TH01 e o vWA. Em 1998, por iniciativa da *International Criminal Police Organization* (Interpol), foram adicionados dois novos marcadores: o FGA e o D21S11. Passado um ano, em 1999, o conjunto de marcadores a ser utilizado no registo da Interpol teve um acréscimo de três marcadores, contando nessa altura com os marcadores TH01, vWA, FGA, D21S11, D3S1358, D8S1179 e D18S51 (Martin et al. 2001).

A nova lista de sete marcadores foi igualmente publicada na resolução do conselho de 25 de Junho de 2001 da União Europeia relativa ao intercâmbio de resultados de análises de ADN que considerou “conveniente estabelecer uma primeira lista mínima dos marcadores de ADN utilizados nas análises de ADN para fins judiciais nos Estados-Membros e que possa ser utilizada no âmbito desse intercâmbio” (União Europeia 2001).

Desde 1997 que nos Estados Unidos da América eram utilizados treze marcadores e passados poucos anos concluiu-se que os sete marcadores utilizados na Europa não eram suficientes. Deste modo, em 2009, foi publicada uma nova resolução do conselho relativa ao intercâmbio de resultados de análises de ADN que considerou “que o intercâmbio de dados de ADN entre os Estados-Membros se encontra rapidamente a progredir e que as bases de dados nacionais de ADN estão a aumentar em dimensão e número, e recordando que o valor estatístico dos dados de ADN corresponde à probabilidade de concordância aleatória e depende inteiramente do número de marcadores de ADN que foram analisados de forma segura, considera-se que é necessário aumentar a atual Série Normalizada Europeia de Loci, aprovada em 2001”. Deste modo, foram acrescentados cinco marcadores. Atualmente, o ESS conta com doze marcadores: TH01, vWA, FGA, D21S11, D3S1358, D8S1179, D18S51 D12S391, D1S1656, D10S1248, D22S1045 e D2S441 (União Europeia 2009).

Apesar do ESS e do CODIS serem mencionados como dois conjuntos distintos de marcadores utilizados em dois continentes diferentes, também será possível uma eventual troca intercontinental de resultados uma vez que ambos os sistemas têm sete marcadores em comum, já para não referir os marcadores extra utilizados na rotina dos laboratórios forenses

que poderá corresponder a alguns dos *loci* STR presentes no conjunto de sistemas obrigatório do outro continente. Provavelmente é o que acontece, tendo em conta que os laboratórios usam *kits multiplex* comerciais elaborados exclusivamente para a prática forense, oferecendo normalmente quinze marcadores a serem analisados que constam em ambos CODIS e ESS.

Passados treze anos da escolha inicial dos STRs como marcadores obrigatórios nos laboratórios forenses pelo EDNAP, estes continuam a ser os marcadores de eleição para a análise de ADN, o intercâmbio de resultados e a criação de bases de dados.

2.2.3. Kits comerciais

Atualmente são utilizados *kits multiplex* comerciais para amplificação do ADN, sendo estes compostos por:

- ✓ mistura de *primers* para a PCR desenhados com vista à amplificação de um determinado conjunto de *loci* STR;
- ✓ tampão contendo reagentes necessários à reação;
- ✓ polimerase de ADN;
- ✓ *ladder* alélico com os alelos mais comuns para os *loci* dos STRs a serem amplificados para permitir a identificação do tamanho de repetição dos alelos;
- ✓ controlo positivo de ADN para verificar o correto funcionamento de todos os reagentes do *kit* (Butler 2012).

No Laboratório do Serviço de Genética e Biologia Forenses (SGBF) do INMLCF,I.P. são normalmente utilizados dois *kits multiplex* comerciais manuseados por técnicos diferentes,

sendo desta forma possível confirmar no final o resultado dos marcadores em comum nos dois *kits*.

Na amplificação das amostras são utilizados os *kits* comerciais das famílias: *AmpFlSTR® Identifiler®* e *PowerPlex® HS System*. Ambos os *kits* têm em comum treze marcadores autossômicos incluídos no sistema CODIS: CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, FGA, TH01, TPOX e vWA, para além do marcador de género amelogenina (Budowle 1998). É de referir que estes *kits* são desenvolvidos especificamente tendo em consideração as necessidades dos laboratórios forenses.

- AmpFlSTR® Identifiler® Plus, Applied Biosystems

Além dos treze STRs tetranucleotídicos incluídos no sistema CODIS, possui mais dois marcadores: D2S1338 e D19S433.

- PowerPlex® 16 HS System, Promega

Além dos treze STRs tetranucleotídicos incluídos no sistema CODIS possui mais dois marcadores pentanucleotídicos: Penta D e Penta E

“Se descobrirem um suspeito, ele pode ser relacionado com o homicídio através do ADN. Naturalmente que também estamos a procurar na nossa base de dados, mas é raro conseguirmos encontrar algo por essa via. Por enquanto, o registo ainda é demasiado pequeno. Apenas podemos sonhar com o dia em que teremos o ADN de todos os cidadãos numa base de dados para investigação.”

Gritos do Passado, Camilla Lackberg 2004

2.2.4. Bases de dados de perfis genéticos

Em Abril de 1995 o Reino Unido sagrou-se o primeiro país do mundo a ter uma base de dados genéticos nacional, contando atualmente com cerca de 5 milhões de amostras (Aavv 2009). Posteriormente, foram muitos os países que seguiram as pisadas do Reino Unido, incluindo Portugal.

Após treze anos, foi criada no nosso país a Lei n.º 5/2008 de 12 de Fevereiro que no Artigo 1º “estabelece os princípios de criação e manutenção de uma base de dados de perfis de ADN, para fins de identificação, e regula a recolha, tratamento e conservação de amostras de células humanas, a respetiva análise e obtenção de perfis de ADN, a metodologia de comparação de perfis de ADN, extraídos das amostras, bem como o tratamento e conservação da respetiva informação em ficheiro informático.” Esta base serve única e exclusivamente para fins de identificação civil e criminal, evidenciando-se uma preocupação frequente com a salvaguarda da privacidade (Artigo 23º, 27º, 28º, 33º, 35º e 36º) e o direito à informação (Artigo 9º e 24º). Apesar do perfil genético ser obtido a partir das regiões não-codificantes do ADN, existe uma certa apreensão relacionada com a privacidade genética uma vez que a análise do ADN permite a determinação da predisposição para certas patologias tal como o seu diagnóstico e

tratamento (Pinheiro 2008). Um dos motivos pelo qual não existem bases de dados com amostras de toda a população prende-se exatamente com razões de privacidade e questões éticas. Ainda que alvo de algumas controvérsias, a Islândia tornou-se o primeiro (e ainda único) país a possuir uma base de dados genéticos de toda a população, apesar do objetivo principal desta base estar relacionado com a descoberta de novos métodos para diagnóstico, prevenção e cura de algumas doenças (Gulcher 2000).

Independentemente de toda esta incessante preocupação com a salvaguarda da vida privada e tendo em conta as vantagens e desvantagens na existência de uma base de dados com o perfil genético de toda a população, reconhece-se que a utilidade e os benefícios excedem copiosamente os inconvenientes. Assim, os perfis genéticos presentes na base de dados de Portugal pertencem a voluntários, profissionais que trabalham diretamente com o ADN (seja na recolha ou análise), condenados de crimes dolosos com pena de prisão igual ou superior a três anos e amostras problema para identificação civil ou criminal (Artigo 15º).

Um vestígio desconhecido deixado num local de crime, um corpo esqueletizado, vítimas de um desastre em massa, um vestígio de uma pessoa desaparecida e um recém-nascido abandonado. Todas estas situações têm algo em comum: ADN. E todo este ADN partilha algo em comum: individualidade. E, sem embargo, de que serve a individualização nestes casos, se não houver nada com que se possa fazer a comparação? Um vestígio num local de crime para o qual não há suspeitos. Um corpo esqueletizado, e nenhuma pessoa dada como desaparecida. Vítimas de um desastre de massa sem famílias que reclamem os corpos. O vestígio de uma pessoa desaparecida de quem só o nome se sabe. E, no fim, um bebé recém-nascido abandonado numa aldeia transmontana onde nenhuma mulher grávida foi vista nos últimos nove meses. Todos estes casos poderiam custar milhares de euros ao Estado durante a

investigação e, mesmo assim, nunca apresentarem um resultado positivo. A base de dados genéticos pode coadjuvar, definitivamente, a resolução de todas estas situações.

A título de exemplo da utilidade e necessidade das bases de dados de perfis genéticos é de salientar que a Interpol apoia fortemente a criação de bases de dados de ADN, com o máximo número de perfis possíveis. Nas recomendações publicadas em Janeiro de 2015 pelo *DNA monitoring expert group* pertencente à Interpol, salientam-se as vantagens das bases de dados no auxílio e reforço da Lei:

- No combate a crimes violentos como homicídios e agressões;
- No combate a crimes em grande escala como assaltos e roubo de viaturas;
- No auxílio à identificação de pessoas desaparecidas e restos cadavéricos não identificados;
- No estabelecimento de uma ferramenta poderosa e de baixo custo no combate à criminalidade;
- Na identificação de possíveis autores de crimes bem como a ligação entre diferentes cenas de crime;
- Na exclusão de indivíduos de um inquérito e na exoneração de condenados inocentes;
- No combate a crimes internacionais como o tráfico de estupefacientes, o tráfico de pessoas e o terrorismo.

No Reino Unido, entre 2008 e 2009, aproximadamente 6 em 10 perfis genéticos encontrados em locais de crime, tiveram uma correspondência na NDNAD (*National DNA Database*) (Aavv 2009). E isto, apenas com 10% da população britânica presente na NDNAD. As probabilidades de correspondência entre perfis genéticos se toda a população constasse da

base de dados seria extremamente elevada poupando muito tempo e recursos, permitindo a resolução de investigações forenses de uma forma mais breve e efetiva.

2.3. Genética Populacional e Estatística em Genética Forense

A grande vantagem do uso do ADN para identificação genética, não reside, *per se*, no facto de estar estatisticamente provado que não existem dois indivíduos com o mesmo ADN. O grande valor do uso desta molécula, e tratando-se toda esta problemática com assuntos relacionados com a legislação e o Direito, reside nos valores estatisticamente significativos e nas elevadas probabilidades que o ADN oferece, sendo atualmente aceite como prova nos tribunais portugueses (Lei n.º 45/2004, de 19 de Agosto).

Atenta-se, assim, que por detrás do valor destas probabilidades estão cálculos com as frequências alélicas. E que frequências são estas? Poderiam ser utilizadas as frequências alélicas dos Estados Unidos da América calculadas por um dos mais famosos geneticistas forenses como o Dr. John Butler ou, então, utilizar as frequências alélicas do país mais próximo. Ambas as respostas estão incorretas.

2.3.1. Genética Populacional

Uma determinada população é caracterizada pelo seu painel particular de frequências alélicas relativamente a dadas características. Cada país tem uma população distinta e apesar de biologicamente não existirem fronteiras legais, uma população escandinava não se assemelha a uma população queniana em certos aspetos. No entanto, a população do norte de Portugal poderá ser mais semelhante à população da Galiza, do que à população do Algarve, dada a distância geográfica. Em países com uma área geográfica muito extensa, como é o caso do Brasil ou da China, existem vários estudos das frequências alélicas exclusivamente para

determinadas regiões desses países (Huang et al. 2013; Hessab et al. 2015). Também em Portugal, que apesar de ser um país com uma área geográfica relativamente pequena, existem estudos para o norte, centro, sul e arquipélagos do país (Vieira-Silva 2006; Lopes 2009; Pontes 2012).

Em relação aos marcadores para os quais as frequências são estimadas, estes não podem ser escolhidos de forma aleatória, mas antes escolher as frequências alélicas dos *loci* STRs dos *kits* utilizados para a amplificação das amostras nesse laboratório.

Assim, as frequências alélicas a utilizar pelo Laboratório do Serviço de Genética e Biologia Forenses da Delegação do Sul do Instituto Nacional de Medicina Legal e Ciências Forenses, I.P. (INMLCF,I.P.) são as frequências pertencentes à população abrangida pela Delegação do Sul do INMLCF,I.P. referente aos marcadores presentes nos *kits AmpFlSTR® Identifier® Plus* e *PowerPlex® 16 HS System* utilizados pelo laboratório: CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, FGA,TH01, TPOX, vWA, D2S1338, D19S433, Penta D e Penta E.

Base de dados de frequências alélicas

Após o conhecimento da população a que se cinge o estudo, importa recolher os dados de uma amostra significativa dessa população de modo a poder caracteriza-la com base na informação obtida. Neste caso, importa ter dados de indivíduos de todos os distritos abrangidos pela Delegação do Sul.

O tamanho da amostra é outro fator a ter em conta. De acordo com algumas comparações efetuadas entre uma centena de amostras e várias centenas de amostras foram obtidos

resultados similares para as frequências alélicas (Butler 2010). No entanto, quanto maior for a amostra, maior será a representatividade da população. Não é aconselhável efetuar extrapolações de estudos de 100 indivíduos para uma população com dezenas de milhares de pessoas. E ainda, quanto maior for o número de amostras maior será a probabilidade de encontrar alelos raros, podendo ser feito o cálculo da frequência também para estes alelos menos comuns.

Após a reunião de todos os perfis genéticos da população de interesse, são calculadas as frequências alélicas através da contagem direta dos alelos e realizados testes estatísticos de interesse forense. Existem vários *softwares* que fazem estes cálculos como o *Arlequin*, o *PowerStats* e o *Genepop* a título de exemplo (Raymond 1995; Tereba 1999; Excoffier 2010).

Equilíbrio de Hardy-Weinberg

Os genes são herdados de acordo com as Leis de Mendel. Cada indivíduo herda um alelo da mãe e outro do pai, de modo aleatório. O Equilíbrio de Hardy-Weinberg (HWE) demonstra a tendência para a manutenção de dadas características genéticas ao longo das gerações, em determinadas condições. Assim, e de acordo com o HWE, as frequências alélicas devem permanecer constantes de uma geração para a seguinte em condições ideais (Hardy 1908; Weinberg 1908). O HWE aplica-se a populações infinitamente grandes e panmíticas, onde não ocorrem mutações, não se registam fenómenos de seleção natural nem acontecimentos de migração. Se existir alguma fuga ao HWE também será possível demonstrar que o equilíbrio volta a ser restabelecido ao fim de alguns cruzamentos ao acaso. Desta forma não são introduzidos novos alelos e, conseqüentemente, as frequências alélicas permanecem

constantes. O teste do HWE serve para determinar se os alelos num determinado *locus* são independentes e se há excesso de homozigotia (Butler 2010). Contudo, na natureza, em populações reais, estes parâmetros não são mantidos, até porque as populações humanas não se tratam de populações fechadas, pelo que as frequências se podem alterar ao longo do tempo, originando uma eventual evolução da mesma. A introdução de novos alelos, a ocorrência de mutações e a seleção natural são os grandes motores da evolução de uma espécie.

Para um determinado marcador com dois alelos, seja p e q a representação das frequências desses dois alelos. Assim, a frequência esperada de um genótipo homozigótico será p^2 ou q^2 , tratando-se do alelo p ou q , respetivamente. A frequência do genótipo heterozigótico será pq . O HWE pode ser demonstrado matematicamente pelo binómio $(p+q)^2$ que se desenvolve na expressão $p^2+2pq+q^2$ que corresponde à distribuição genotípica que se prevê encontrar numa população infinitamente grande, com cruzamentos aleatórios, sem seleção, mutação ou migração.

Índice de fixação

Geralmente também são feitos estudos comparativos da distância genética entre algumas populações distintas através do estudo do índice de fixação (F_{ST}). Este é uma medida da estruturação em populações naturais que tem em conta a variação genética e a diferença nas frequências alélicas entre as populações que compara. O valor de F_{ST} pode variar entre 0 e 1. Quando o valor corresponde a 1, significa que existe um isolamento entre as populações, isto é, não existe partilha de alelos. Logo, quanto maior for a distância genética entre as

populações, maior será o valor do F_{ST} e vice-versa. Todavia, é de referir que tratando-se de populações humanas, o F_{ST} nunca tomará o valor de 1, já que as comparações são feitas dentro de populações da mesma espécie (Wright 1950; Reynolds 1983; Slatkin 1995).

2.3.2. Parâmetros estatísticos de interesse forense

Alguns dos parâmetros estatísticos de interesse forense que podemos considerar são a heterozigotia (H), o poder de discriminação (PD), o poder de exclusão (PE), o conteúdo de informação polimórfica (PIC), a frequência do alelo mínimo (MAF) e o índice de paternidade típico (TPI).

Heterozigotia

A heterozigotia (H) é uma medida de variabilidade genética que corresponde à proporção de indivíduos heterozigóticos na população. É calculada dividindo o número de amostras que contêm alelos heterozigóticos pelo número total de amostras. A heterozigotia observada (H_o) corresponde à proporção de indivíduos heterozigóticos encontrada no total das amostras. Já a heterozigotia esperada (H_e) corresponde à heterozigotia observada, considerando que a população se encontra em HWE e calcula-se através da fórmula

$$H_e = 1 - \sum_i^n p_i^2 ,$$

em que p_i é a frequência do $i^{\text{ésimo}}$ alelo numa população de n amostras (Butler 2005). Quanto maior for a heterozigotia maior será a diversidade alélica e, conseqüentemente, menor será a probabilidade de uma amostra aleatória ser igual a outra.

Poder de discriminação

O poder de discriminação (PD) ou probabilidade de discriminação corresponde ao potencial poder de diferenciar dois indivíduos escolhidos aleatoriamente da população, ou seja, que estes apresentem genótipos diferentes. O PD obtém-se através da aplicação da fórmula

$$PD = 1 - 2 \left(\sum_{i=1}^n p_i^2 \right)^2 - \sum_{i=1}^n p_i^4$$

e transmite-nos o quão discriminativos são os *loci* em estudo em termos de individualização (Butler 2005).

Poder de exclusão

O poder de exclusão (PE) ou probabilidade de exclusão consiste na probabilidade que um determinado marcador genético tem para excluir um genótipo em particular, ou seja, excluir um determinado indivíduo. Por outras palavras, o PE pode ser definido como a fração de indivíduos cujo perfil de DNA é distinto do perfil genético de um indivíduo escolhido aleatoriamente da população (Huston 1998). O PE pode ser determinado através da fórmula

$$PE = H^2 (1 - (1 - H)H^2) ,$$

em que H corresponde à heterozigotia (Butler 2005).

Conteúdo de informação polimórfica

Os vários tipos de marcadores genéticos diferem entre si pelo conteúdo informativo que apresentam, o qual depende do seu grau de polimorfismo. O conteúdo de informação polimórfica, *polymorphic information content* (PIC), é um indicador da qualidade de um dado marcador genético, proporcionando uma estimação do poder de discriminação do mesmo (Botstein 1980; Nagy 2012). Para que um marcador seja considerado informativo, o valor de PIC deverá ser sempre superior a 0,5. O PIC pode ser determinado através da fórmula

$$PIC = 1 - \sum_1^n p_i^2 - \left(\sum_1^n p_i^2 \right)^2 + \sum_1^n p_i^4$$

Frequência do alelo mínimo

Como já foi mencionado, de momento é impossível ter uma representação dos genótipos de toda a população, pelo que é necessário selecionar uma amostra representativa da população. Desta forma, pode ocorrer uma lacuna de representatividade de todos os alelos raros presentes na população em estudo. O relatório do *National Research Council* de 1996 (NRC II) estipula que a frequência alélica calculada para um alelo tão raro que aparece apenas um par de vezes não é fidedigna. Assim, recomenda-se que cada alelo deverá ser observado pelo menos cinco vezes para que a sua frequência seja utilizada em cálculos estatísticos. Com o

fim de contornar esta situação foi introduzida uma forma de cálculo da frequência do alelo mínimo, ou *minimum allele frequency* (MAF), representada por

$$MAF = \frac{5}{2N}$$

em que N corresponde ao número total de indivíduos, sendo calculada para cada marcador separadamente (National Research Council Committee on DNA Forensic Science 1996). Consequentemente, sempre que a frequência de um alelo for inferior à MAF, durante a realização dos cálculos estatísticos será utilizado o valor obtido após a aplicação da fórmula $5/2N$ para o marcador em que o alelo é observado. A MAF é totalmente dependente e influenciada pelo tamanho da amostra, sendo-lhe inversamente proporcional. Ou seja, quanto maior for o tamanho da amostra, menor será o valor da MAF.

Índice típico de paternidade

O índice de paternidade típico, *typical paternity index* (TPI), é um cálculo que tem como objetivo demonstrar, para um dado marcador genético, quantas vezes é mais provável que o indivíduo testado seja o pai biológico em comparação com um homem escolhido aleatoriamente da população. Valores de TPI inferiores a 1 são indicativos de ausência de parentesco. O índice de paternidade calcula-se através da fórmula

$$PI = \frac{H + h}{2h}$$

em que h corresponde à homozigotia (Huston 1998; Butler 2005).

2.3.3. Valor estatístico do ADN

Quando nos deparamos com uma amostra problema em contexto criminal e existe um suspeito, entram em jogo duas hipóteses: exclusão ou não exclusão. Numa exclusão o perfil genético do suspeito não coincide com o da amostra problema, ao passo que na existência de não exclusão existe uma correspondência entre ambos. É de referir que uma investigação também poderá ser inconclusiva devido a uma vastidão de fatores, não podendo ser feitas assunções em respeito à correspondência ou à falta dela entre amostras e suspeitos.

Nos casos em que ocorre exclusão torna-se fácil garantir que aquele determinado indivíduo não está relacionado com o crime em questão, pelo menos não o está ao nível da contribuição com o seu material genético. No entanto, se o perfil genético da amostra problema e do suspeito coincidirem surge um vasto número de questões do tipo de assunções que poderão ser feitas. Será sempre certa a correspondência entre o perfil de uma amostra problema e o perfil de um suspeito? E no caso de existirem dois ou mais indivíduos com o mesmo perfil genético para os marcadores analisados? Sabemos que o ADN individualiza todos os seres vivos, individualizando inclusive os seres dentro da mesma espécie, tornando cada membro de uma população único. Mas dada a inexequibilidade da análise do ADN na íntegra, apenas são analisados alguns marcadores genéticos. E assim, algures no passado poderia ter existido um indivíduo com o mesmo perfil genético para esses marcadores. Como atualmente, no presente, também pode existir. A única forma de solucionar este problema consiste na aplicação de cálculos estatísticos com resultados escritos em probabilidades que dão voz à interpretação da correspondência entre os perfis genéticos da amostra e do suspeito. Urge, ainda, referir que uma probabilidade elevada não é, *per se*, indicação de que o suspeito perpetróu o crime, pelo que devem ser tomados com prudência outros aspetos pelo julgador.

As probabilidades foram concebidas de modo a atribuir um valor numérico entre 0 e 1 a assuntos em que não há certezas absolutas, sendo 0 o acontecimento impossível e 1 o acontecimento certo. Neste caso, e tratando-se de ciência a probabilidade equivalente a 1 é uma impossibilidade.

Para uma situação de não-exclusão, qual será então a probabilidade da amostra problema pertencer ao suspeito?

Likelihood ratio

O *likelihood ratio* (LR), ou razão de verosimilhança, consiste na comparação de probabilidades sob a forma de duas hipóteses diferentes, sendo estas mutuamente exclusivas, ou seja, sendo só uma delas possível.

Neste caso, uma das hipóteses assume que a amostra problema pertence ao suspeito e a outra hipótese assume o oposto, ou seja, a amostra pertence a um indivíduo ao acaso da população que, coincidentemente, apresenta o mesmo perfil genético que o suspeito (Butler 2010). Assim,

$$LR = H1/H2$$

Onde H1 corresponde à hipótese da amostra pertencer ao suspeito e H2 corresponde à hipótese da amostra pertencer a um indivíduo ao acaso da mesma população.

Dada a correspondência genética entre amostra e suspeito, H1 assumirá o valor de 1. Em relação a H2, este valor será calculado tendo em conta as frequências alélicas para cada *locus* analisado e se o indivíduo é homozigótico ou heterozigótico para determinado *locus*. Se for

homozigótico a frequência alélica é elevada ao quadrado (p^2), se for heterozigótico multiplicam-se ambas as frequências entre si e por 2 ($2pq$) (Butler 2010).

Relativamente às investigações de averiguação da paternidade, e não podendo o pretenso pai ser excluído, é feito o cálculo da probabilidade do pretenso pai ser, efetivamente, o pai biológico. Para tal é calculado o Índice de Paternidade (IP) que, através de uma divisão matemática, considera a existência de duas probabilidades condicionais: a hipótese de o pretenso pai ser o pai biológico e a hipótese de o pai biológico ser um homem aleatório da mesma população. Nesta divisão a primeira hipótese ocupa o lugar do numerador e a segunda hipótese será o denominador. O IP trata-se, então, de uma razão de verosimilhança (*likelihood ratio*) entre duas probabilidades que competem com diferentes hipóteses, refletindo, deste modo, quantas vezes é mais provável que o pretenso pai tenha o conjunto de alelos necessários do que um homem ao acaso da mesma população (Butler 2005). Assim,

$$IP=X/Y,$$

onde X é a probabilidade do pretenso pai ser o pai biológico e Y é a probabilidade de um homem ao acaso da mesma população ser o pai biológico.

Quando os cruzamentos ocorrem de forma aleatória, a probabilidade de Y, ou seja, que este indivíduo ao acaso transmita um alelo específico será equivalente à frequência desse alelo na população em causa. Mais uma vez denota-se a grande utilidade em manter as frequências genéticas atualizadas.

O valor de X pode ser 0,5 ou 1, no caso do pretenso pai ser heterozigótico ou homozigótico, respetivamente, para o alelo em causa. O IP é calculado separadamente para cada marcador,

sendo no final multiplicados estes valores obtendo o índice de paternidade combinado para o conjunto de todos os *loci* analisados.

Como se denota com base nestes exemplos, as frequências alélicas são essenciais para toda a análise estatística que culmina, inevitavelmente, com as conclusões da investigação.

3.Objetivos do estudo

Com este trabalho pretendeu-se fazer a atualização das frequências alélicas e dos principais parâmetros estatísticos de interesse populacional e forense dos 17 *loci* STR autossómicos analisados na casuística forense, presentes nos dois *kits* comerciais *AmpFlSTR® Identifier® Plus* e *PowerPlex® 16 HS System* utilizados na rotina do laboratório do Serviço de Genética e Biologia Forenses da Delegação do Sul do INMLCF,I.P. Além da determinação das frequências alélicas, foram calculados os principais parâmetros estatísticos de interesse populacional e forense: os valores de homozigotia e de heterozigotia, o poder de discriminação e de exclusão, o equilíbrio de Hardy-Weinberg (valor de p), o conteúdo de informação polimórfica, a frequência do alelo mínimo e o índice típico de paternidade. No final foi realizada a comparação entre os dados obtidos neste estudo e no estudo anterior para a mesma população (Ribeiro et al. 2006) e ainda o cálculo do índice de fixação e avaliação das distâncias genéticas entre a população portuguesa e sete populações oriundas de distintas zonas geográficas, nomeadamente: Espanha, Itália, Grécia, Roménia, Marrocos, Angola e Coreia.

Este tipo de trabalho é tão importante que vários estudos são realizados todos os anos por vários laboratórios forenses nacionais e internacionais. No entanto, estes parâmetros populacionais podem sofrer alterações pelo que a sua atualização é de grande relevância para a prática forense, nomeadamente para a prática do laboratório forense onde o trabalho prático desta tese foi desenvolvido.

Uma vez que a cada ano que passa a população portuguesa sofre mudanças demográficas e tendo em conta que o último estudo realizado sobre as frequências alélicas da região abrangida pela Delegação do Sul do INMLCF,I.P. efetuou-se há quase dez anos, em 2006 (Vieira-Silva 2006), tornava-se imprescindível atualizar estes dados. Para tal foram utilizados os perfis genéticos de 5362 indivíduos portugueses caucasianos envolvidos em investigações

de parentesco entre os anos de 2005 a 2014. Para garantir a ausência de parentesco foram utilizadas única e exclusivamente as amostras dos pretensos pais e das mães (ou pretensas mães em investigações de averiguação de maternidade).

Este estudo, ao ser realizado na Delegação do Sul do INMLCF, I.P., englobou amostras de indivíduos provenientes dos distritos de Santarém, Portalegre, Lisboa, Setúbal, Évora, Beja e Faro. O somatório da área total destes distritos compreende quase metade do país continental, residindo aqui cerca de 42,1% do total da população portuguesa (Instituto Nacional de Estatística 2012), o qual se espera ser representativo da população estudada de modo a poder ser usado na casuística forense.

4.Material e métodos

4.1. Rotina laboratorial

A partir do momento que uma amostra biológica chega ao laboratório, é necessário executar a extração e amplificação do ADN até à obtenção do perfil genético dessa amostra. Tratando-se o presente trabalho de um estudo retrospectivo que abrange um número elevado de amostras e tendo em conta que estas foram extraídas e amplificadas no passado, sendo já conhecido o perfil genético das mesmas, não foi necessário efetuar o normal percurso laboratorial. Ainda assim, torna-se relevante clarificar de forma sucinta o percurso das amostras biológicas provenientes de investigações de parentesco no Laboratório de Genética e Biologia Forenses da Delegação do Sul do INMLCF, I.P. desde a extração até à obtenção do perfil genético das amostras.

Extração

O primeiro passo consiste na extração do ADN a partir das amostras biológicas de sangue e saliva (epitélio bucal) recolhidas sob forma de mancha de sangue ou em zaragatoa bucal, colhidas com consentimento informado. Neste laboratório é utilizado, por rotina, o método de extração com Chelex 100[®] (Walsh 1991) para a extração de amostras de referência.

Amplificação pela PCR

A reação em cadeia da polimerase trata-se de um processo enzimático automatizado no qual uma determinada região de interesse do ADN é replicada de forma artificial milhares de vezes. Assim, no final desta reação são obtidas um número elevado das sequências de ADN,

tantas quantas forem necessárias. No SGBF da Delegação do Sul são utilizados por rotina os *kits* de PCR *Identifiler*[®] *Plus* (Applied Biosystems) e *Powerplex*[®] *16 HS* (Promega).

Eletroforese capilar

Para a separação e detecção dos alelos dos STRs é utilizado o método da eletroforese capilar. O equipamento de eletroforese capilar Applied Biosystems 3130xl Genetic Analyzer encontra-se ligado a um computador com os *softwares* necessários para a detecção dos STRs. Após a corrida da amostra, é possível visualizar o perfil genético correspondente através do *software* *GeneMapper*[®] (Applied Biosystems).

4.2.Amostragem e compilação dos dados

A população de interesse

Todas as semanas são realizadas dezenas de investigações de averiguação de parentesco com algum dos intervenientes natural ou residente em qualquer um dos distritos abrangidos pela Delegação do Sul do INMLCF,IP.: Santarém, Portalegre, Lisboa, Setúbal, Évora, Beja e Faro. Existe uma grande heterogeneidade a nível da proveniência das amostras, desde portugueses até à comunidade imigrante ou mesmo a intervenientes que nem sequer residam ou sejam naturais de Portugal.

A grande maioria dos indivíduos envolvidos nestas investigações são Portugueses caucasianos residentes na área abrangida pela Delegação do Sul. No entanto, há uma percentagem considerável de indivíduos pertencentes às comunidades africanas e brasileira, e, mais recentemente, um número crescente da comunidade chinesa, envolvidos em investigações de parentesco.

Para este estudo, a população de interesse corresponde a todos os portugueses caucasianos sem qualquer relação de parentesco entre si.

A compilação dos dados

Para o trabalho desta dissertação, começou por ser feita a passagem manual de todos os perfis genéticos com as informações relativas ao grupo étnico, naturalidade e parentesco desde os anos de 2005 até 2014 para uma folha de cálculo Microsoft Office Excel 2007, para poder excluir no final todos os perfis que não correspondessem aos requisitos necessários, ou seja, Portugueses caucasianos não aparentados residentes na área selecionada intervenientes em investigações de averiguação da paternidade/maternidade.

Após a seleção dos perfis genéticos de interesse, foi atribuído um número a cada perfil, começando no 1. Os dados relativos ao grupo étnico, parentesco e naturalidade foram eliminados, garantindo deste modo o anonimato completo dos dados genéticos. Posteriormente à exclusão das amostras que não se enquadravam nos requisitos, restaram 5362 perfis genéticos. O perfil genético destas amostras pode ser consultado no Anexo disponível eletronicamente.

Todos os dados foram confirmados e completados utilizando os dados em bruto armazenados na base de dados do *software GeneMapper* que foram transferidos eletronicamente para a folha de cálculo de Excel a partir dos ficheiros do sequenciador, além dos dados que foram introduzidos manualmente, pelo que foi realizada uma validação cruzada de todos os dados, quer dos que foram inseridos eletronicamente e dos que foram inseridos manualmente.

Após toda a compilação e validação dos dados, foi utilizada a folha de *Excel PowerStats* modificada de acordo a poder abarcar todos os dados usados, devido ao elevado número de amostras. Nesta folha foi efetuado o cálculo para as frequências alélicas, o PD, o PE, a Ho, o TPI e o PIC (Tereba 1999).

Para o cálculo do HWE (valores de p), da H_e e para a construção da matriz com as distâncias genéticas interpopulacionais foi utilizado o *software Arlequin v.3.5* (Excoffier & Lischer 2010). Este *software* também realiza o cálculo das frequências alélicas, pelo que estas foram calculadas novamente utilizando o *Arlequin* para verificar a fidedignidade dos valores obtidos através da folha de cálculo *PowerStats*. Todas as frequências provaram estar em concordância.

O desenho da árvore filogenética foi obtido de acordo com o algoritmo *Neighbor-Joining* através do programa *Molecular Evolutionary Genetics Analysis* (MEGA) v.6.0. (Tamura 2013).

5.Resultados e Discussão

5.1. Frequências alélicas

Na Tabela 1 estão representadas as frequências alélicas e os parâmetros estatísticos de interesse forense para os 17 *loci* STRs estudados.

Como se pode observar, estamos perante resultados extremamente ricos na quantidade de alelos para os *loci* estudados, o que advém do elevado número de amostras incluídas. Este estudo conta, na totalidade, com 81 alelos diferentes para o total dos 17 *loci* STRs, sendo este um número muito elevado para estudos do mesmo género. Por exemplo, e como meio de comparação, no estudo português sobre a mesma população e para os mesmos *loci*, havia um total de 66 alelos (Vieira-Silva 2006). Para o mesmo número de loci, na China há um total de 31 alelos (Li 2009). No Líbano o número equivale a 53 alelos (Chouery 2010). Em Cabo Verde existe um total de 43 alelos (Vieira-Silva 2008). Em Angola o número total é de 53 alelos (Beleza 2004). Possivelmente, isto deve-se à elevada amostragem, já que o número total de amostras excede os cinco milhares.

Tabela 1: Frequências alélicas e parâmetros estatísticos para a população do Sul de Portugal

Alelo	D3S1358	TH01	D21S11	D18S51	PentaE	D5S818	D13S317	D7S820	D16S539
2.2									
3		0,0001							
3.2									
4		0,0007							
4.2					0,0002				
5		0,0009			0,0620		0,0001		
6		0,2079			0,0012		0,0002		0,0001
6.3								0,0005	
7		0,1716			0,1418	0,0004	0,0001	0,0169	
7.3									
8		0,1334			0,0179	0,0086	0,1361	0,1552	0,0237
8.3		0,0003							
9		0,1968		0,0002	0,0132	0,0308	0,0563	0,1286	0,1199
9.1								0,0003	
9.3		0,2753							
10	0,0001	0,0131		0,0137	0,0852	0,0611	0,0495	0,2717	0,0595
10.1								0,0001	
10.2				0,0005					
10.3		0,0001							
11	0,0011			0,0091	0,1311	0,3405	0,3275	0,2274	0,3042
11.2									
12	0,0027			0,1402	0,2032	0,3767	0,2772	0,1623	0,2834
12.1						0,0002			0,0001
12.2									
13	0,0029			0,1302	0,1127	0,1710	0,1115	0,0311	0,1787
13.2				0,0001					
14	0,1049			0,1429	0,0571	0,0104	0,0405	0,0054	0,0291
14.2				0,0002					
14.4					0,0001				
15	0,2645			0,1467	0,0454	0,0005	0,0009	0,0005	0,0012
15.2				0,0001					
15.4					0,0001				
16	0,2509			0,1585	0,0410				0,0001
16.1									
16.2									
17	0,2082			0,1121	0,0363				0,0001
17.2									
18	0,1528			0,0651	0,0184				
18.1				0,0001					
18.2									
18.3	0,0001								
19	0,0113			0,0421	0,0174				
19.2									
20	0,0005			0,0242	0,0081				
21				0,0088	0,0054				
21.2									

CSF1PO	PentaD	vWA	D8S1179	TPOX	FGA	D2S1338	D19S433	Alelo
	0,0085							2.2
								3
	0,0006							3.2
								4
								4.2
	0,0012			0,0002				5
	0,0004			0,0045				6
								6.3
0,0026	0,0050		0,0001	0,0059				7
0,0001								7.3
0,0082	0,0192		0,0105	0,5110				8
								8.3
0,0231	0,1948		0,0104	0,1083			0,0004	9
								9.1
								9.3
0,2671	0,1169		0,0864	0,0641			0,0017	10
								10.1
								10.2
0,0001								10.3
0,3182	0,1623	0,0005	0,0974	0,2751			0,0059	11
							0,0002	11.2
0,3188	0,1896	0,0002	0,1259	0,0300			0,1055	12
							0,0001	12.1
							0,0019	12.2
0,0519	0,1981	0,0027	0,2887	0,0008		0,0002	0,2664	13
							0,0149	13.2
0,0090	0,0764	0,1016	0,2322	0,0002		0,0004	0,3175	14
							0,0325	14.2
								14.4
0,0009	0,0214	0,1225	0,1163	0,0001		0,0022	0,1410	15
							0,0469	15.2
								15.4
	0,0042	0,2357	0,0282		0,0002	0,0440	0,0433	16
					0,0001			16.1
							0,0164	16.2
	0,0012	0,2640	0,0034		0,0024	0,2613	0,0032	17
							0,0015	17.2
	0,0003	0,1835	0,0006		0,0138	0,0771	0,0001	18
								18.1
					0,0006		0,0009	18.2
								18.3
		0,0735			0,0656	0,1052		19
					0,0002			19.2
		0,0138			0,1393	0,1427		20
		0,0020			0,1809	0,0358		21
					0,0032			21.2

Alelo	D3S1358	TH01	D21S11	D18S51	PentaE	D5S818	D13S317	D7S820	D16S539
22				0,0039	0,0013				
22.1									
22.2									
23				0,0011	0,0007				
23.2									
24				0,0004	0,0002				
24.1									
24.2			0,0027						
24.3									
25			0,0008	0,0001					
25.2			0,0011						
26			0,0007						
26.2			0,0002						
27			0,0235						
28			0,1470						
28.2			0,0001						
29			0,2299						
29.2			0,0008						
29.3			0,0003						
30			0,2441						
30.2			0,0367						
31			0,0601						
31.2			0,1111						
32			0,0097						
32.2			0,0919						
33			0,0008						
33.1			0,0002						
33.2			0,0311						
34			0,0004						
34.2			0,0048						
35			0,0013						
35.1			0,0001						
35.2			0,0009						
36			0,0001						
44.2									
N	10724	10724	10724	10724	10578	10724	10724	10724	10724
Ho	0,7870	0,7910	0,8450	0,8700	0,8810	0,7000	0,7750	0,7970	0,7700
He	0,7890	0,7950	0,8390	0,8770	0,8880	0,7080	0,7780	0,8060	0,7760
HWE	0,5837	0,1908	0,0393	0,0004	0,7263	0,1930	0,1963	0,7022	0,3560
MP	0,0770	0,0740	0,0450	0,0280	0,0230	0,1340	0,0820	0,0650	0,0830
PD	0,9230	0,9260	0,9550	0,9720	0,9770	0,8660	0,9180	0,9350	0,9170
PIC	0,7560	0,7630	0,8200	0,8640	0,8800	0,6570	0,7460	0,7780	0,7420
PE	0,5740	0,5820	0,6860	0,7350	0,7560	0,4280	0,5530	0,5930	0,5440
TPI	2,344	2,387	3,234	3,858	4,180	1,664	2,221	2,462	2,173
MAF	0,000466	0,000466	0,000466	0,000466	0,000473	0,000466	0,000466	0,000466	0,000466

CSF1PO	PentaD	vWA	D8S1179	TPOX	FGA	D2S1338	D19S433	Alelo
					0,1709	0,0359		22
					0,0001			22.1
					0,0062			22.2
					0,1513	0,1010		23
					0,0035			23.2
					0,1371	0,1025		24
					0,0001			24.1
					0,0007			24.2
					0,0001			24.3
					0,0777	0,0790		25
					0,0001			25.2
					0,0342	0,0118		26
								26.2
					0,0079	0,0010		27
					0,0027	0,0001		28
								28.2
					0,0008			29
					0,0001			29.2
								29.3
					0,0001			30
								30.2
								31
					0,0001			31.2
								32
					0,0002			32.2
								33
								33.1
								33.2
								34
								34.2
								35
								35.1
								35.2
								36
					0,0001			44.2
10724	10580	10724	10724	10724	10724	10702	10704	N
0,7190	0,8320	0,8010	0,8140	0,6420	0,8670	0,8500	0,7920	Ho
0,7230	0,8400	0,8100	0,8160	0,6470	0,8650	0,8630	0,7920	He
0,0538	0,0006	0,0978	0,4290	0,0000	0,1153	0,4464	0,3134	HWE
0,1280	0,0460	0,0620	0,0570	0,1770	0,0330	0,0320	0,0710	MP
0,8720	0,9540	0,9380	0,9430	0,8230	0,9670	0,9680	0,9290	PD
0,6710	0,8200	0,7840	0,7920	0,5950	0,8500	0,8490	0,7640	PIC
0,4590	0,6600	0,6010	0,6260	0,3440	0,7280	0,6950	0,5840	PE
1781	2,975	2,515	2,692	1,396	3,755	3,336	2,402	TPI
0,000466	0,000473	0,000466	0,000466	0,000466	0,000466	0,000467	0,000467	MAF

Legenda da Tabela 1 - N: Número de alelos; Ho: Heterozigotia observada; He: Heterozigotia esperada; HWE: Equilíbrio de Hardy-Weinberg; MP: Probabilidade de *matching*; PD: Poder de discriminação; PIC: Conteúdo de informação polimórfica; PE: Poder de exclusão; TPI: Índice de paternidade típico; MAF: Frequência do alelo mínimo

5.2. Parâmetros estatísticos de interesse forense

Os marcadores D18S51, Penta E, FGA e D2S1338 apresentam-se como os mais variáveis do conjunto, com 16 ou mais alelos cada e uma heterozigotia esperada acima dos 85%.

O marcador TPOX mostrou ser o menos polimórfico, apresentando os menores valores de H , de PD , de PE , de PIC e de TPI . Este marcador apresenta o valor mais elevado de probabilidade de *matching*.

O marcador Penta E salienta-se como o mais polimórfico com os valores mais elevados de H , de PD , de PE , de PIC e de TPI . Este marcador apresenta o valor mais baixo de probabilidade de *matching*.

Relativamente ao HWE, denotam-se quatro marcadores com um desvio estatisticamente significativo ao equilíbrio, com valores de p inferiores a 0,05. Após a aplicação da correção de Bonferroni, através da fórmula $0.05/17$, o valor de p toma o valor de 0,0029, restando três marcadores com desvio ao equilíbrio: D18S51, Penta D e TPOX. Existem três motivos principais para a ocorrência deste desvio (Butler 2005):

- Relação de parentesco entre os progenitores, o que leva a um aumento da homozigotia;
- Subestrutura populacional, o que leva igualmente a um aumento da homozigotia devido a cruzamentos não-aleatórios;
- Seleção devido à sobrevivência de indivíduos com diferentes genótipos que se reproduzem a diferentes taxas.

O marcador com o valor de p mais baixo é também o marcador que apresenta o menor valor de heterozigotia. O facto do marcador TPOX apresentar o valor de 0, não invalida os dados do

estudo, até porque de todos os marcadores do CODIS, este provou ser o menos variável entre indivíduos (Butler 2010). A título de exemplo, existem alguns estudos publicados onde o valor de p também iguala 0, nomeadamente os artigos “*U.S. population data for 29 autosomal STR loci*” (Hill 2013) e “*Population database of 17 autosomal STR loci from the four predominant Eastern Slovakia regions*” (Soták 2011).

5.3. Comparação interpopulacional

Na Tabela 2 é possível observar os valores de F_{ST} entre a população do sul de Portugal e as populações de Marrocos (Bentayebi 2014), Angola (Melo 2010), Roménia (Anghel 2014), Itália (Brisighelli 2009), Espanha (García 2012), Grécia (Sánchez-Diz 2008) e Coreia (Hong 2013).

Tabela 2: Matriz de distâncias genéticas entre o sul de Portugal e sete países distintos.

	Marrocos	Angola	Roménia	Itália	Espanha	Grécia	Sul de Portugal	Coreia
Marrocos	x							
Angola	0,00500	x						
Roménia	-0,00056	0,01142	x					
Itália	0,00046	0,01511	-0,00102	x				
Espanha	0,00262	0,01857	-0,00167	-0,00184	x			
Grécia	-0,00256	0,00893	-0,00165	-0,00141	0,00009	x		
Portugal	0,00028	0,01106	-0,00112	0,00032	-0,00007	-0,00082	x	
Coreia	0,01588	0,00952	0,01718	0,02718	0,02848	0,02009	0,01967	x

A Figura 2 representa a árvore filogenética com base nos valores dos índices de F_{ST} .

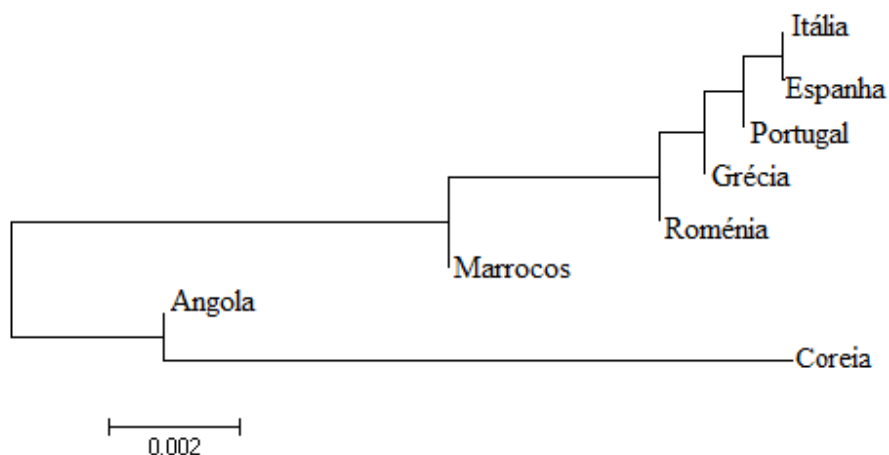


Figura 2: Relações filogenéticas entre as populações do Sul de Portugal, Espanha, Itália, Grécia, Roménia, Marrocos, Angola e Coreia.

Com base nos valores de F_{ST} constata-se que as populações geneticamente mais distintas de Portugal são Angola e Coreia, com valores de F_{ST} de 0,01106 e 0,01967, respetivamente. Essa distância também é visível na árvore filogenética. As populações do sul da Europa, nomeadamente: Itália, Espanha e Grécia são as mais próximas de Portugal, como seria de esperar.

5.4. Comparação com o estudo de 2006

Ambos os estudos utilizaram os mesmos marcadores e população da mesma área, mas em intervalos de tempo distintos.

Em termos gerais não são observáveis alterações expressivas nas frequências alélicas de ambos os estudos, apesar do estudo atual ser muito mais rico na quantidade de alelos que apresenta no total. Isto deve-se possivelmente ao número elevado de amostras utilizadas.

Relativamente aos os parâmetros estatísticos de interesse forense, na H_o , na H_e e no PD os valores obtidos são semelhantes para ambos os estudos. Os valores do PE são semelhantes à exceção do marcador TPOX, onde o PD é igual a 0,4143 no estudo de 2006, diferindo de 0,3440 no estudo atual, apesar de não ser uma diferença muito significativa. Em relação ao HWE, é de notar que os valores de p apresentam algumas diferenças.

6. Conclusões

Este estudo cumpriu o seu objetivo principal que consistiu na atualização das frequências alélicas e parâmetros forenses utilizados pelo Laboratório de Genética e Biologia Forenses instalado na Delegação do Sul do INMLCF,I.P.

Através deste trabalho foi possível concluir que:

- ✓ as frequências alélicas podem sofrer alterações, apesar de neste caso essas alterações não terem sido muito expressivas, até porque estamos a falar de um período de cerca de dez anos desde o estudo anterior. Durante um intervalo de dez anos para uma população humana dificilmente haveria uma mudança brusca nas frequências alélicas;
- ✓ o aumento do tamanho da amostra condiciona alguns resultados, nomeadamente a observação de alelos raros que não estavam presentes no estudo anterior. Para o conjunto dos 17 marcadores, foi observado um total de 81 alelos a contrastar com um total de 66 para o estudo de 2006;
- ✓ é sempre importante manter as frequências alélicas e os parâmetros estatísticos atualizados, principalmente na prática forense já que uma probabilidade de identidade elevada poderá ter implicações jurídicas com forte impacto na vida das pessoas em causa.

7.Referências bibliográficas

- Aavv, 2009. *The National DNA Database - Annual report 2007-2009*,
- Anghel, A. et al., 2014. Genetic polymorphism data on 15 autosomal STR markers in a Western Romanian population sample. *Legal medicine (Tokyo, Japan)*, 16(4), pp.238–40.
- Beleza, S. et al., 2004. 17 STR data (AmpF/STR Identifiler and Powerplex 16 System) from Cabinda (Angola). *Forensic Science International*, 141(2-3), pp.193–196.
- Bentayebi, K. et al., 2014. Genetic ancestry of a Moroccan population as inferred from autosomal STRs. *Meta gene*, 2, pp.427–38.
- Botstein, D. et al., 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American journal of human genetics*, 32(3), pp.314–331.
- Brisighelli, F. et al., 2009. Allele frequencies of fifteen STRs in a representative sample of the Italian population. *Forensic Science International: Genetics*, 3(2).
- Budowle, B., Moretti, T.R., et al., 1998. CODIS and PCR-Based Short Tandem Repeat Loci: Law Enforcement Tools. *Second European Symposium on Human Identification*, pp.73–88.
- Budowle, B., Moretti, T.R., et al., 1998. *CODIS and PCR-based short tandem repeat loci: law enforcement tools. Proceedings of the Second European Symposium on Human Identification*, Madison, WI.
- Busam, K.J., 2010. *Dermatopathology, Foundations in Diagnostic Pathology Series*, Elsevier.
- Butler, J.M., 2012. *Advanced Topics in Forensic DNA Typing: Methodology*, San Diego, CA: Elsevier.
- Butler, J.M., 2005. *Forensic DNA typing: biology, technology, and genetics of STR markers*, Amsterdam: Elsevier.
- Butler, J.M., 2010. *Fundamentals of Forensic DNA Typing*, Amsterdam: Elsevier.
- Butler, J.M., 2006. Genetics and genomics of core short tandem repeat loci used in human identity testing. *Journal of Forensic Sciences*, 51(2), pp.253–265.
- Calabuig, J.A.G. & Cañadas, E.V., 2004. *Medicina legal y toxicología*. Barcelona: Masson.
- Chouery, E. et al., 2010. Population genetic data for 17 STR markers from Lebanon. *Legal Medicine*, 12(6), pp.324–326.
- Dettmeyer, R., Verhoff, M.A. & Schütz, H.F., 2014. *Forensic Medicine: Fundamentals and Perspectives*. Berlin: Springer.

- Edwards, A. et al., 1991. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *American journal of human genetics*, 49(4), pp.746–56.
- Edwards, A. et al., 1992. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics*, 12(2), pp.241–253.
- Europeia, U., 2001. Council Resolution of 25 June 2001 on the exchange of DNA analysis results. *Official Journal of European Communities*, pp.1–4.
- Excoffier, L. & Lischer, H.E.L., 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10(3), pp.564–567.
- García, O. et al., 2012. Population genetic data and concordance study for the kits Identifiler, NGM, PowerPlex ESX 17 System and Investigator ESSplex in Spain. *Forensic science international. Genetics*, 6(2), pp.e78–9.
- Geserick, G. & Wirth, I., 2012. Genetic Kinship Investigation from Blood Groups to DNA Markers. *Transfusion medicine and hemotherapy : offizielles Organ der Deutschen Gesellschaft für Transfusionsmedizin und Immunhamatologie*, 39(3), pp.163–175.
- Gill, P. et al., 1994. Identification of the remains of the Romanov family by DNA analysis. *Nature genetics*, 6(2), pp.130–5.
- Gill, P., Jeffreys, A. & Werrett, D., 1985. Forensic application of DNA “fingerprints.” *Nature*, 318(12), pp.577–579.
- Gulcher, J.R. & Stefánsson, K., 2000. The Icelandic Healthcare Database and informed consent. *The New England journal of medicine*, 342(24), pp.1827–30.
- Haak, W. et al., 2010. Ancient DNA from European early neolithic farmers reveals their near eastern affinities. *PLoS biology*, 8(11), p.e1000536.
- Hardy, G.H., 1908. MENDELIAN PROPORTIONS IN A MIXED POPULATION. *Science (New York, N.Y.)*, 28(706), pp.49–50.
- Hessab, T. et al., 2015. Genetic data on 17 STR autosomal loci for a sample population of the State of Rio de Janeiro, Brazil. *Forensic science international. Genetics*, 14, pp.e4–5.
- Hill, C.R. et al., 2013. U.S. population data for 29 autosomal STR loci. *Forensic science international. Genetics*, 7(3), pp.e82–3.
- Hong, S.B. et al., 2013. Korean population genetic data and concordance for the PowerPlex® ESX 17, AmpFISTR Identifiler®, and PowerPlex® 16 systems. *Forensic Science International: Genetics*, 7(3).

- Huang, Q. et al., 2013. Genetic polymorphism of 15 STR loci in Chinese Han population from Shanghai municipality in East China. *Forensic science international. Genetics*, 7(2), pp.e31–4.
- Huston, K.A., 1998. Statistical Analysis of STR Data. , pp.14–15.
- Instituto Nacional de Estatística, 2012. Resultados Definitivos. *Censos 2011*, 26, pp.1–41.
- Jackson, A. & Jackson, J., 2011. *Forensic Science* 3rd ed., Harlow: Pearson.
- Jeffreys, A.J., 1979. DNA sequence variants in the G gamma-, A gamma-, delta- and beta-globin genes of man. *Cell*, 18(1), pp.1–10.
- Jeffreys, A.J., Wilson, V. & Thein, S.L., 1985. Individual-specific “fingerprints” of human DNA. *Nature*, 316(6023), pp.76–9.
- Li, C. et al., 2009. Genetic polymorphism of 17 STR loci for forensic use in Chinese population from Shanghai in East China. *Forensic Science International: Genetics*, 3(4).
- Lopes, V. et al., 2009. Allelic frequency distribution of 17 STRs from Identifiler and PowerPlex-16 in Central Portugal area and the Azores archipelago. *Forensic science international. Genetics*, 4(1), pp.e1–7.
- Martin, P.D., Schmitter, H. & Schneider, P.M., 2001. A brief history of the formation of DNA databases in forensic science within Europe. *Forensic Science International*, 119(2), pp.225–231.
- Melo, M.M. et al., 2010. Genetic study of 15 STRs loci of Identifiler system in Angola population. *Forensic Science International: Genetics*, 4(5).
- Nagy, S. et al., 2012. PICcalc: an online program to calculate polymorphic information content for molecular genetic studies. *Biochemical genetics*, 50(9-10), pp.670–2.
- National Research Council Committee on DNA Forensic Science, 1996. *The Evaluation of Forensic DNA Evidence*,
- Pinheiro, M.F.T., 2008. *CSI Criminal*, Porto: Universidade Fernando Pessoa.
- Pontes, M.L. & Pinheiro, M.F., 2012. Population data of the AmpFISTR® NGM™ STR loci in a North of Portugal sample. *Forensic science international. Genetics*, 6(5), pp.e127–8.
- Raymond, M. & Rousset, F., 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, 86, pp.248–249.
- Reynolds, J., Weir, B.S. & Cockerham, C.C., 1983. Estimation of the coancestry coefficient: Basis for a short-term genetic distance. *Genetics*, 105(3), pp.767–779.

- Ribeiro, T. et al., 2006. South Portugal population genetic analysis with 17 loci STRs. *International Congress Series*, 1288, pp.367–368.
- Sánchez-Diz, P. et al., 2008. 16 STR data of a Greek population. *Forensic Science International: Genetics*, 2(4).
- Der Sarkissian, C. et al., 2013. Ancient DNA reveals prehistoric gene-flow from siberia in the complex human population history of North East Europe. *PLoS genetics*, 9(2), p.e1003296.
- Schmitt, A., Cunha, E. & Pinheiro, J., 2006. *Forensic anthropology and medicine: Complementary sciences from recovery to cause of death*. Totowa, N.J.: Humana Press.
- Slatkin, M., 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, 139(1), pp.457–462.
- Soták, M. et al., 2011. Population database of 17 autosomal STR loci from the four predominant Eastern Slovakia regions. *Forensic science international. Genetics*, 5(3), pp.262–3.
- Tamura, K. et al., 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, 30(12), pp.2725–2729.
- Tereba, A., 1999. Tools for analysis of population statistics. *Profiles in DNA*, pp.14–16.
- Tilstone, W.J., Savage, K.A. & Clark, L.A., 2006. *Forensic Science: An Encyclopedia of History, Methods, and Techniques*, ABC-CLIO.
- União Europeia, 2009. Council Resolution of 30 November 2009 on the exchange of DNA Analysis Results. *Official Journal of European Communities*, pp.1–3.
- Vieira-Silva, C. et al., 2008. Cabo Verde population study with 17 STR loci. *Forensic Science International: Genetics Supplement Series*, 1(1), pp.380–382.
- Vieira-Silva, C. et al., 2006. South Portugal population genetic analysis with 17 loci STRs. *International Congress Series*, 1288, pp.367–368.
- Walsh, P.S., Metzger, D.A. & Higuchi, R., 1991. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques*, 10, pp.506–513.
- Weinberg, W., 1908. Über den Nachweis der Vererbung beim Menschen. *Jahreshefte des Vereins für Vaterländische Naturkunde in Württemberg*, 64, pp.369–382.
- Wong, Z. et al., 1987. Characterization of a panel of highly variable minisatellites cloned from human DNA. *Annals of human genetics*, 51(Pt 4), pp.269–288.
- Wright, S., 1950. Genetical structure of populations. *Nature*, 166(4215), pp.247–249.

8. Anexos

Artigo submetido à revista Forensic Science International: Genetics

FSIGEN-S-15-00244

Title: Analysis of 17 STR data on 5362 South Portuguese individuals - an update on reference database

Article Type: Correspondence.

First Author: Raquel Cabezas Silva.

